# SUPERVISED LEARNING METHOD

# KNN ALGORITHM: GLASS DATASET

This algorithm is a supervised learning algorithm, where the destination is known, but the path to the destination is not. Understanding nearest neighbors forms the quintessence of machine learning. The items present in the groups are homogeneous in nature. Now, suppose we have an unlabeled example which needs to be classified into one of the several labeled groups. How do you do that? Unhesitatingly, using KNN Algorithm.

Choosing the number of nearest neighbors i.e. determining the value of k plays a significant role in determining the efficacy of the model. Thus, selection of k will determine how well the data can be utilized to generalize the results of the KNN algorithm. A large k value has benefits which include reducing the variance due to the noisy data; the side effect being developing a bias due to which the learner tends to ignore the smaller patterns which may have useful insights.

**Algorithm:**

* Calculate the distance from x to all points in your data.
* Sort the points in your data by increasing distance from x.
* Predict the majority label of the k closest points.

Note that the value of k effects the results, its ideal to test the model for different values of k for better results and there by a better model.

## Data Introduction:

**Glass Identification Database** from UCI contains 10 attributes including id. The response is glass type which has 7 discrete values.

### **Attributes:**

### **Id**: 1 to 214 (removed from CSV file) **RI**: refractive index **Na**: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10) **Mg**: Magnesium **Al**: Aluminum **Si**: Silicon **K**: Potassium **Ca**: Calcium **Ba**: Barium **Fe**: Iron **Type of glass**: (Class Attribute)

### 1- building\_windows\_float\_processed 2 - building\_windows\_non\_float\_processed 3 - vehicle\_windows\_float\_processed 4 - vehicle\_windows\_non\_float\_processed (none in this database) 5 – containers 6 – tableware 7 - headlamps

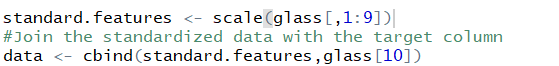
## Data Reading:

The data was imported from UCI Machine learning repository to glass.



## Standardize the Data:

It is important to standardize the data in KNN algorithm. Here, I will standardize the data using scale() function. I excluded the target column Type while scaling.



## Data Summary:

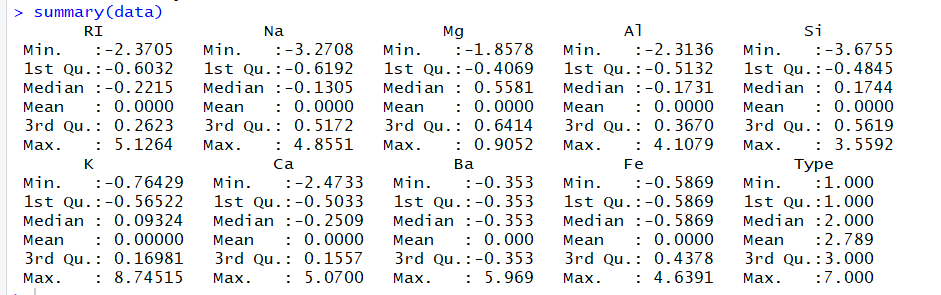
First, we check for any missing value.

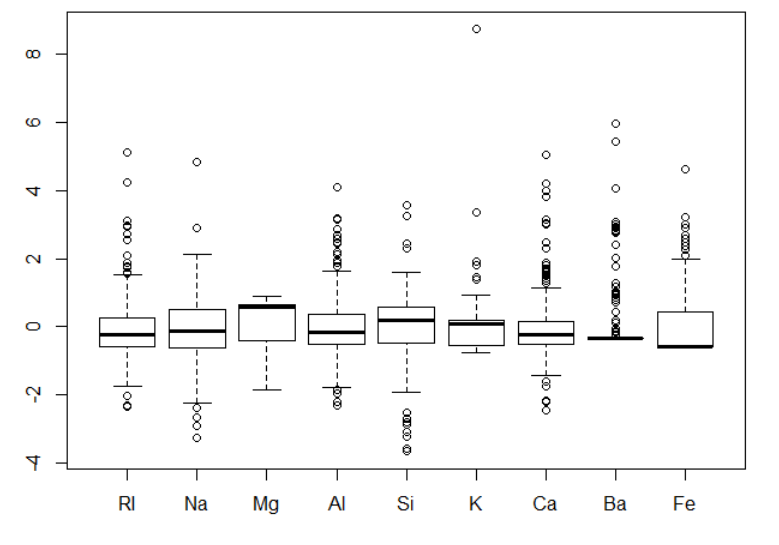


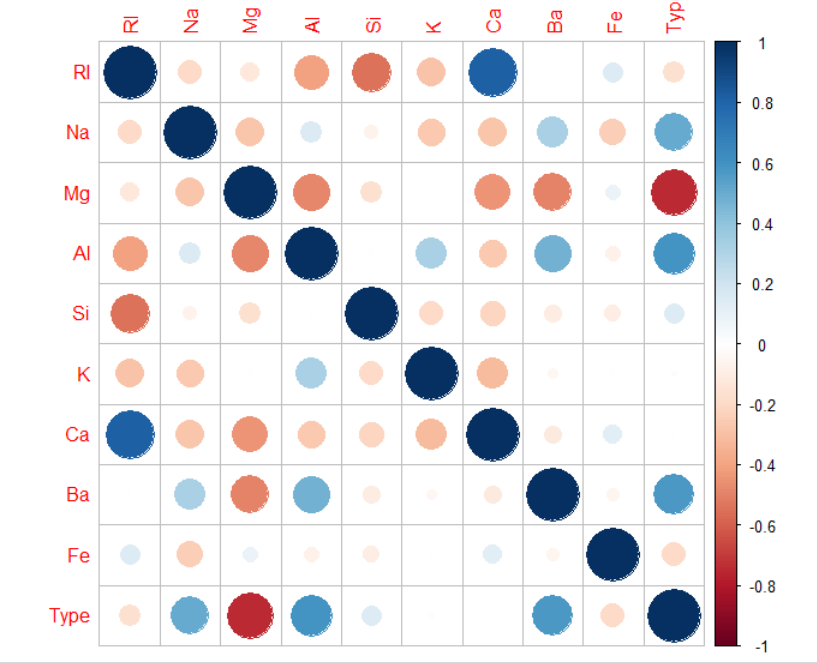
A sample data looks like:

## 

Summary of the data are below:



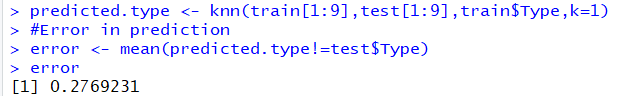




## Modeling using KNN:

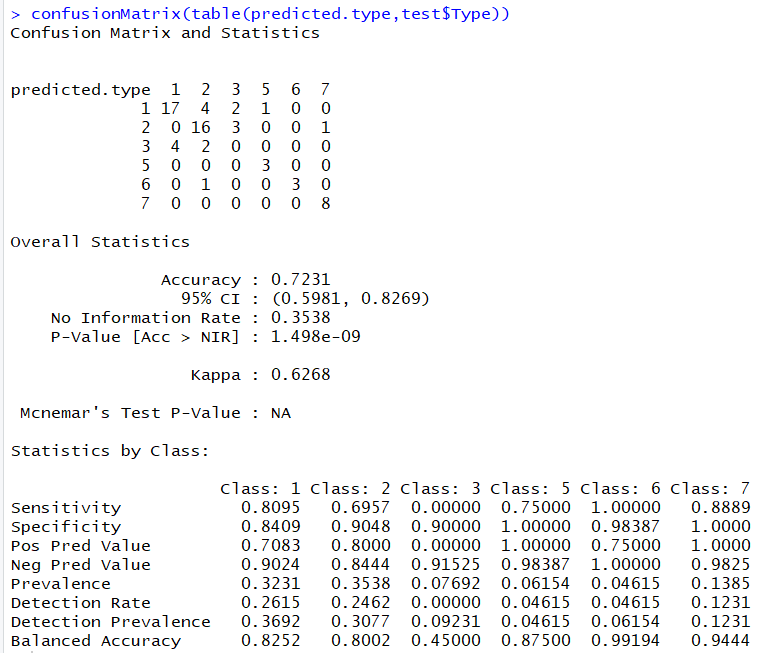
A seed value was set using set.seed(). I have used catools() to split the data into training and testing in the ratio of 70/30.

A KNN model is fit using knn() function to predict the target type of the test data set and using k=1.



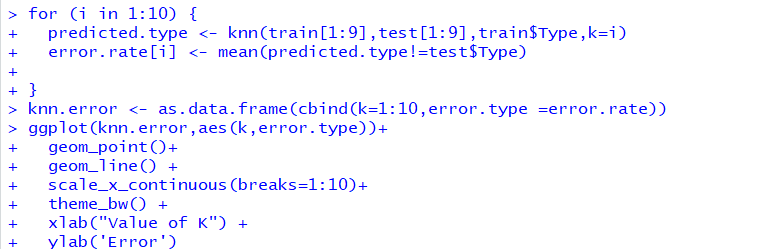
The error rate for this model was 27.69% and accuracy of 72.31%

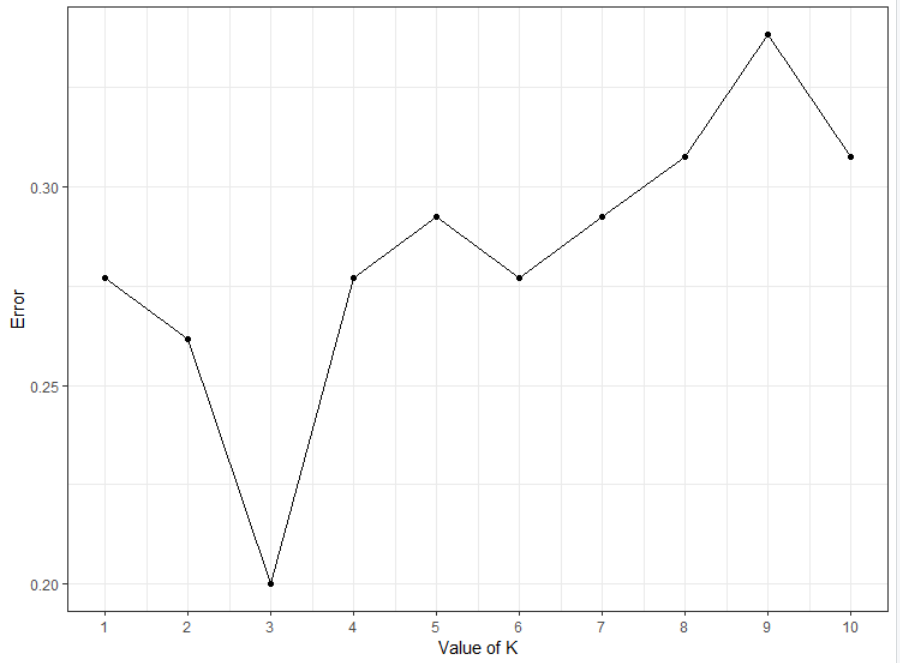
Below is the confusion matrix for the model.

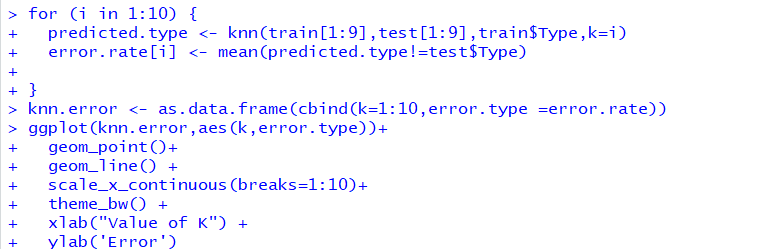


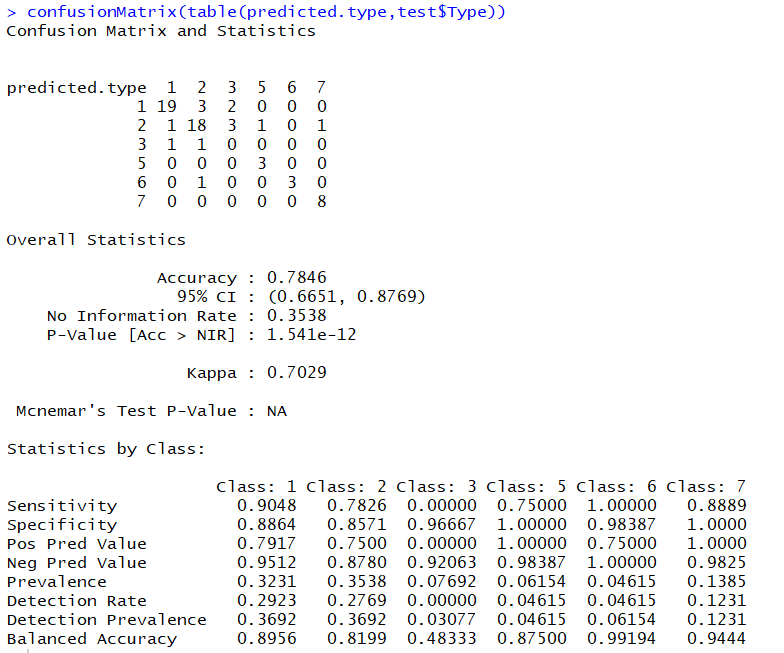
## Choosing K value:

To find the optimal value of k, I plotted a graph between different values of k and the error rate. The value of k with minimum error in the below graph is at k=3.



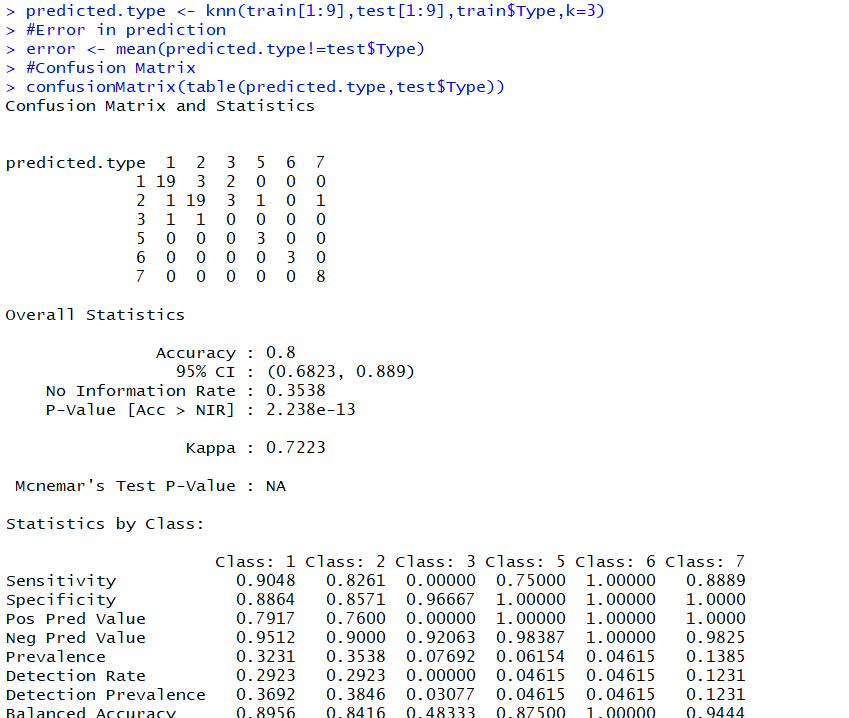






## Final Model:

Now, I will build the final model using k=3





The accuracy of this final model was found to be 80% with 20% error rate.

## Conclusion:

The accuracy achieved for classifying the different glasses using KNN algorithm was 80% with an error rate 20%.